Complete and assembled genome sequences of Pantoea calida DSM22759T and Pantoea gaviniae DSM22758T

Stevens, Marc J A ; Zurfluh, Katrin ; Stephan, Roger

Abstract: The genomes of Pantoea calida DSM 22759T and Pantoea gaviniae DSM 22758T were sequenced using single-molecule real-time sequencing. They consist of a 4.3-Mbp chromosome containing 4,092 genes, of which 3,977 encode proteins, and a 4.5-Mbp chromosome containing 4,236 genes, of which 4,120 encode proteins, respectively.

DOI: https://doi.org/10.1128/genomeA.00157-18

Complete and assembled genome sequences of *Pantoea calida* DSM22759^T and *Pantoea gaviniae* DSM22758^T

Marc J. A. Stevens, Katrin Zurfluh, Roger Stephan[#]

Institute for Food Safety and Hygiene, Vetsuisse Faculty, University of Zurich, Switzerland.

[#]Address correspondence Roger Stephan: roger.stephan@uhz.ch

Abstract. The genomes of *Pantoea calida* DSM22759^T and *Pantoea gaviniae* DSM22758^T were sequenced using single molecule real time sequencing. They consist of a 4.3-Mbp chromosome containing 4092 genes of which 3977 encode for proteins and a 4.5 Mbp chromosome containing 4236 genes of which 4120 encode for proteins, respectively.

The genus *Pantoea* belongs to the order of *Enterobacteriales* and is, since the taxonomic update of the family *Enterobacteriaceae*, placed in the family *Erwiniaceae* (1). The two species *Pantoea gaviniae* and *Pantoea calida* form a distinct clade within the genus *Pantoea*, are closely related (average nucleotide identity of 88 %), but have an average nucleotide identity of only 73.44-78.48 % with other members of the genus (2). Here the completely sequenced and assembled genomes of the type strains *Pantoea calida*

DSM22759^T and *Pantoea gaviniae* DSM22758^T, originally isolated from infant formula and an infant formula production environment, are presented.

DNA was isolated using the Wizard® Genomic DNAPurification Kit (Promega AG, Dübendorf, Switzerland). Sequencing was performed using single molecule real time sequencing on 2 PacBio RS II cells (Pacific Biosciences, Menlo Park, CA, USA) at the Functional Genomics Center Zurich, Switzerland. Raw reads were filtered using the RS Filter Only protocol in the SMRT-portal (Pacific Biosciences) with 0.85 as polymerase read quality cut off and a minimal length of 500 bp. For *Pantoea calida* DSM22759^T, a total of 81,593 reads with an average length of 15,916 bp passed filtering, corresponding to 1,298,688,643 sequenced base pairs and a coverage of approximately 300-fold. The reads were assembled using the Canu assembler (3) with the option "pacbio-raw" and an estimated genome size of 4.4 Mbp. The Canu output consisted of two contigs, the first containing the chromosomal sequence and the second a phage sequence, as identified using PHAST (4). The phage sequence was not submitted to Genbank. The chromosomal contig was polished in CLC workbench 7 (CLC, Aarhus, Denmark) and is 4308453 bp long. The G+C content of the chromosome is 57.0 %. The start of the chromosome was identified using OriC (5) and the genome was annotated by the NCBI Prokaryotic Genomes Automatic Annotation Pipeline server. The genome of *Pantoea calida* DSM22759^T contains 4092 genes of which 3977 encode for proteins. Further 83 tRNA genes and 7 ribosomal RNA operons are present.

For *Pantoea gaviniae* DSM22758^T, a total of 60,962 reads with an average length of 10,288 bp were selected, resulting in 627,151,760 sequenced base pairs and a coverage of approximately 140 fold. The reads were assembled using the Canu assembler 1.6 (3) with

the option "pacbio-raw" to an estimated genome size of 4.5 Mbp. The Canu output consisted of a single contig that was subsequently polished in CLC workbench 7 (CLC, Aarhus, Denmark). The chromosomal origin of replication was identified using OriC-finder 5.0 (5) and the start of the genome set at 9 bp upstream of the first DnaA box in the identified origin region. The genome of *Pantoea gaviniae* DSM22758^T consists of a circular 4,527,605-bp chromosome with a GC content of 58.0 %. The genome was annotated by the NCBI Prokaryotic Genomes Automatic Annotation Pipeline server. The chromosome encodes 4236 genes of which 4120 encode for proteins. Further, 7 ribosomal RNA operons and 86 tRNA genes were identified.

Nucleotide sequence accession number. Sequence and annotation data of the complete genomes of *Pantoea calida* DSM22759^T and *Pantoea gaviniae* DSM22758^T are deposited in the GenBank database with the accession numbers CP026378 and CP026377, respectively.

References

1) Palmer M, Steenkamp ET, Coetzee MPA, Chan WY, van Zyl E, De Maayer P, Coutinho TA, Blom JTH, Duffy B, Vente ST. 2017. Phylogenomic resolution of the

bacterial genus *Pantoea* and its relationship with *Erwinia* and *Tatumella.* Antonie van Leeuwenhoek **110**: 1287

2) Popp A, Cleenwerck I, Iversen C, De Vos P, Stephan R. 2010. *Pantoea gaviniae* sp. nov. and *Pantoea calida* sp. nov., isolated from infant formula and an infant formula production environment. Int J Syst Evol Microbiol **60**:2786-2792.

3) Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res **27**:722-736.

4) Zhou Y. Liang Y, Lynch K, Dennis JJ, Wishart DS. 2011. "PHAST: A Fast Phage Search Tool". Nucl. Acids Res. **39**

5) Gao F, Zhang CT. 2008**.** Ori-Finder: a web-based system for finding oriCs in unannotated bacterial genomes. BMC Bioinformatics **9**:79.